
Generalized Linear Response Method: Application to Hydration Free Energy Calculations

XIN CHEN, ALEXANDER TROPSHA

Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Received 24 November 1997; accepted 16 September 1998

ABSTRACT: A generalized linear response (GLR) method was developed and applied to hydration free energy calculations. According to this method, the atomic hydration can be described as a two-step process. In the first step a point particle is introduced into water, which, according to the scaled particle theory, creates a cavity with the size of a water molecule. The free energy change of this step for the simple point charge (SPC) water model can be calculated as $1.49k_{\text{B}}T$. In the second step the introduced point particle is transformed into a solute atom. The free energy change of this step can be calculated by the linear response approximation, which is applied to van der Waals and electrostatic interactions, as $\langle V_{\text{H}}^a \rangle_{0.5}$. Here V_{H}^a is the solute–water interaction function, and $\langle \cdots \rangle_{0.5}$ denotes the ensemble average at the midpoint of the thermodynamic path between the point particle state and the hydration state. The GLR method was tested by the calculation of hydration free energies of several neutral organic compounds. The results of the calculation were in close agreement with the experiment and were also comparable with those obtained by the conventional free energy simulation method; the computational cost was decreased by about one order of magnitude. The GLR approach is more general than the existing linear response methods for free energy calculations because it applies the linear response approximation to electrostatic and van der Waals interactions and does not incorporate any empirically determined parameters.
© 1999 John Wiley & Sons, Inc. J Comput Chem 20: 749–759, 1999

Keywords: linear response method; free energy calculations; hydration; molecular dynamics simulations

Correspondence to: A. Tropsha; e-mail: tropsha@email.unc.edu

Contract/grant sponsor: Glaxo Wellcome, Inc.

Introduction

Accurate calculation of the free energies associated with many chemical or biochemical processes (solvation, ligand binding, protein folding, etc.) plays an important role in understanding the physical basis of these processes. It is also important in guiding the experimental efforts in the area of rational drug design, especially structure-based drug design. Presently, these calculations are mostly based on the free energy simulation methods, such as slow growth simulations, thermodynamic integration, or free energy perturbations.¹⁻⁴ Unfortunately, their robustness notwithstanding, all these methods are computationally expensive and cannot satisfy the practical requirements of the computational efficiency for the calculations done on large data sets. Recently, several groups considered a faster approach to free

energy calculations based on the linear response approximation.^{5,6} These methods do not require an exhaustive sampling in multiple intermediate states along the transformation path between the initial and final states as in the conventional free energy simulation methods. Therefore, they should provide a more efficient computational route to the accurate free energy calculations.

The most well-known example of a linear response approximation in the treatment of the electrostatic field is provided by Marcus' theory of electron transfer reactions.⁷ According to this theory, an electron transfer system responds linearly to the change of its electrostatic field, which yields the free energy functions with the parabolic form, as illustrated in Figure 1. The reaction coordinate X describes the transformation of the whole system between the initial state A and the final state B , which are defined by their own potential functions V_A and V_B , respectively. X can be some property of the system that depends on the charge

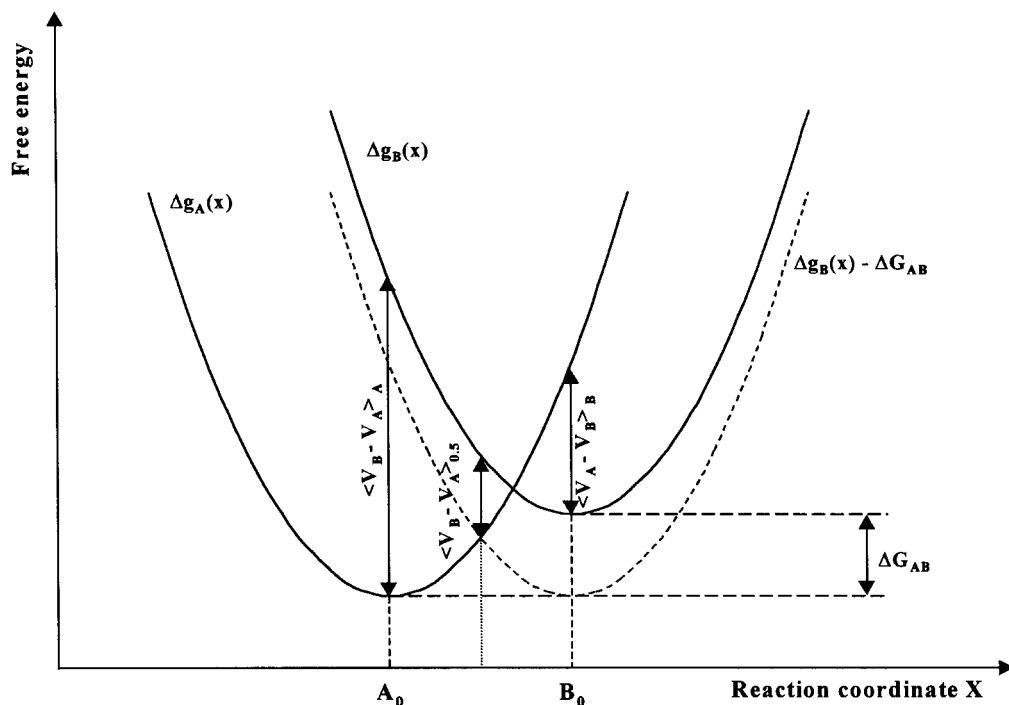


FIGURE 1. Illustration of the free energy curves for a system that obeys the linear response approximation. See the text for the additional description of the solid curves. The vertical dotted line indicates a midpoint between the equilibrium coordinates A_0 and B_0 of the initial and final states, respectively. The dashed curve $[\Delta g_B(x) - \Delta G_{AB}]$ is a copy of the $\Delta g_B(x)$ curve obtained by shifting it downward by ΔG_{AB} . Because $\Delta g_A(x)$ and $[\Delta g_B(x) - \Delta G_{AB}]$ have equal curvature, it is obvious that $\langle V_B - V_A \rangle_{0.5}$ is equal to ΔG_{AB} .

distribution. It is usually assumed that the free energy functions of the system at states A and B can be approximately described by parabolic curves, $\Delta g_A(x)$ and $\Delta g_B(x)$, and that these two parabolic curves have the same force constant (i.e., the whole system responds linearly to the change of electrostatic field with the same dielectric constant). From these postulates it can be easily inferred that the free energy change between the states A and B , ΔG_{AB} , can be expressed as^{5,7}

$$\Delta G_{AB} = (1/2)(\langle V_B - V_A \rangle_A - \langle V_A - V_B \rangle_B), \quad (1)$$

where $\langle \dots \rangle_A$ and $\langle \dots \rangle_B$ denote the ensemble averages in states A and B , respectively. The geometric descriptions of the terms in eq. (1) are shown in Figure 1.

Aqvist et al.⁵ implemented a linear response method in the study of the binding affinities of some endothiapepsin inhibitors, followed by Carlson and Jorgensen⁶ who suggested an extended linear response equation for the calculations of the hydration free energies of several organic compounds. According to Aquvist et al.,⁵ the free energy change of ligand–receptor binding can be calculated from the change of electrostatic and van der Waals interaction energies between the ligand and its environment, because the ligand is transferred from the aqueous solution to the receptor binding site, using the following semiempirical equation:⁵

$$\Delta G_{\text{binding}} = (1/2)\Delta \langle V_{\text{elec}} \rangle + \alpha \Delta \langle V_{\text{vdw}} \rangle, \quad (2)$$

where $\langle V_{\text{elec}} \rangle$ and $\langle V_{\text{vdw}} \rangle$ are the ensemble averages of the electrostatic and van der Waals interaction energies, respectively, between the ligand and its environment, and Δ refers to the difference between these ensemble averages in water and in the receptor binding site. The $1/2$ coefficient before the electrostatic contribution term is derived from the linear response approximation as in eq. (1), and the α coefficient before the van der Waals contribution term is determined by empirical calibration against the experimental data. Aquvist et al.⁵ applied this approach to the calculation of the binding affinities of five structurally different inhibitors of endothiapepsin and found that the α value of 0.162 gave the best agreement with the experimental binding data with a mean unsigned error of ca. 0.3 kcal/mol.

Carlson and Jorgensen⁶ proposed a slightly different version as an extended linear response

equation for the hydration free energy calculations. Their equation includes an additional cavity contribution term and allows the coefficients before all the contribution terms to be fitted. Thus, the hydration free energy, ΔG_{hyd} , can be calculated as⁶

$$\Delta G_{\text{hyd}} = \beta \langle V_{\text{elec}} \rangle + \alpha \langle V_{\text{vdw}} \rangle + \gamma (\text{cavity-term}), \quad (3)$$

where α , β , and γ are the adjustable parameters determined by the fitting to the experimental hydration free energies; and $\langle V_{\text{elec}} \rangle$ and $\langle V_{\text{vdw}} \rangle$ are the ensemble averages of electrostatic and van der Waals interaction energies, respectively, between the compound and water molecules. In their work the cavity contribution term was expressed as one of the three different molecular properties: molecular surface area, solvent accessible surface area, and molecular volume. It was found that if solvent accessible surface area was used, eq. (3) gave the best agreement with the experimental hydration free energies of 14 diverse organic compounds with the root mean square (RMS) deviations of ca. 0.8 kcal/mol for the all-atom model and 0.9 kcal/mol for the united-atom model. More recently, MacDonald and coworkers expanded this set of molecules to 35, giving an RMS deviation from the experimental ΔG_{hyd} of 1.0 kcal/mol and a mean unsigned error of 0.8 kcal/mol.

All previous work on the extension of the linear response approximation to the free energy calculations supported the efficiency of this strategy. However, the empirical parameters, such as α , β , and γ in eqs. (2) and (3), are derived from the analysis of different training sets and, as a result, lack generality. In this article we propose a new generalized linear response (GLR) method for the free energy calculations and derive a new equation for the hydration free energy calculations. This method is based on a special physical model of the hydration process and extends the standard linear response approximation to include the van der Waals interactions. Furthermore, the derived equation does not include any empirical parameters and therefore should be applicable to the hydration free energy calculations of more diverse compounds. We applied this new GLR method in comparison with the conventional slow growth (SG) free energy simulation method to the calculations of absolute hydration free energies of several organic compounds. The results of the GLR method were comparable with those obtained by the SG

method, but the computational cost was decreased by about one order of magnitude.

Methodology

SINGULARITY PROBLEM IN APPLICATION OF LINEAR RESPONSE METHOD TO VAN DER WAALS INTERACTIONS

Free energy calculations are frequently applied to the physicochemical processes characterized by the transfer of a compound between two phases, for example, from a vacuum to aqueous solution (hydration process) or from water to a receptor binding site (binding process). These processes can be theoretically described as the disappearance of a compound from its initial environment followed by its appearance in the final environment. In a physical sense it means that van der Waals and electrostatic interactions of the compound with its initial environment are both eliminated and then established in the final environment. As discussed in the Introduction, the application of the linear response method to calculate the free energy change resulting from the change of electrostatic interaction has been known for many years in the study of the electron transfer process.⁷ However, there have been no reports on the application of the standard linear response method to treat the change of the van der Waals interaction.

From the mathematical viewpoint, the van der Waals and electrostatic potential functions both have a singularity point at zero interaction distance, which may lead to the failure of the linear response approximation as illustrated by the following example. Assume that the state *A* in Figure 1 now represents a system consisting of a solute atom surrounded by the water molecules and state *B* represents this system after the solute atom has disappeared from the water solution. After the van der Waals and electrostatic interactions between the solute atom and the surrounding water have both disappeared in the state *B*, the water molecules have a chance to occupy the former position of the solute atom. If this situation happens, the value of V_A (which still has to be calculated) and therefore $V_A - V_B$ in eq. (1) will be approaching infinity, which obviously makes the linear response approximation impossible to apply (cf. Fig. 1). In fact, such a problem has been encountered in the work of free energy simulations where it was referred to as the "singularity problem." However, if we only allow the electrostatic

interaction between the solute atom and the water to disappear, the presence of the remaining van der Waals interactions will prevent the water molecules from approaching the solute atom too closely and therefore will prevent the singularity problem. This is just the case in the electron transfer processes. These considerations provide a possible explanation as to why the linear response approximation cannot be applied successfully in a straightforward manner to treat the change of the van der Waals interaction.

GLR METHOD FOR HYDRATION FREE ENERGY CALCULATIONS

The new GLR approach avoids the singularity problem by introducing a special step in the calculation pathway as described below. This step allows us to treat the change of the electrostatic and van der Waals interactions by the linear response approximation.

For the GLR method the process of introducing a solute atom into an aqueous solution can be artificially divided into two steps, as illustrated in Figure 2. In the first step a point particle is introduced into water, which, according to the scaled particle theory,^{9,10} creates a cavity of water size. The presence of the point particle in the cavity prevents the singularity problem in the second step where the linear response approximation is used to calculate the free energy change of the transformation of the point particle into the solute atom. The free energies of these two steps are calculated as follows.

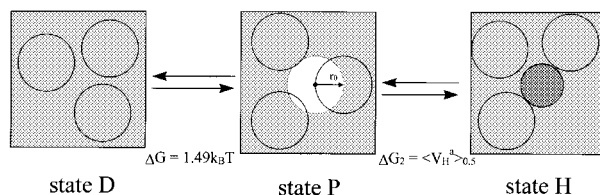


FIGURE 2. The two-step model of the hydration of a single atom. State D is the dehydrated state in which the interactions between the solute atom and water are completely eliminated. State P corresponds to the point particle state in which the interaction between the point particle and water molecules is described by the hard-sphere interaction function as discussed in the text. The white area in the figure indicates the cavity area generated by this point particle. State H corresponds to the hydrated state, which is described by the conventional van der Waals and electrostatic interaction functions.

According to the scaled particle theory,^{9,10} a cavity in the solution can be created by a statistical fluctuation and the probability that it happens, p_0 , is related to the free energy of this process, ΔG_1 , as

$$p_0 = e^{-\Delta G_1/k_B T}, \quad (4)$$

where k_B is the Boltzman constant and T is the temperature of the system. Assuming that the cavity has a spherical shape, the probability of finding the center of a water molecule in this cavity can be calculated as $(4/3)\pi r_0^3 \rho$ (where r_0 is the radius of the cavity and ρ is the number density of water). Because at most only one water molecule can have its center in this region, the probability p_0 can also be expressed as

$$p_0 = 1 - (4/3)\pi r_0^3 \rho. \quad (5)$$

From eqs. (4) and (5) one easily obtains

$$\Delta G_1 = -k_B T \ln(1 - (4/3)\pi r_0^3 \rho). \quad (6)$$

Equation (6) is a direct result of the scaled particle theory and provides a formula to calculate the free energy of introducing a point particle into water. For practical calculations, the value of r_0 was determined as one-half of the equilibrium distance of the water–water van der Waals interaction function. For the simple point charge (SPC) water model, this value is 1.776 Å.¹¹ Substituting $r_0 = 1.776$ Å and $\rho = 0.033$ Å⁻³ (taken from ref. 12) into eq. (6) leads to

$$\Delta G_1 = 1.49 k_B T. \quad (7)$$

In the second step the point particle is transformed into the solute atom. We assume that the free energy change of this process can be calculated by the linear response approximation applied to both van der Waals and electrostatic interactions. The presence of the point particle in the cavity helps to avoid the singularity problem because the water molecules are dispelled outside the cavity. Thus, the point particle in essence plays the same role as the van der Waals envelope in the electron transfer processes, preventing water molecules from approaching the solute atom too closely. Consequently, the free energy change of this step, ΔG_2 , can be calculated using an equation similar to eq. (1), which requires exhaustive sampling in the initial and final states via either molecular dynamics or Monte Carlo simulations.

To simplify the calculations and avoid doing simulations in the point particle state, we devised a slightly different expression to calculate the free

energy change in the second step. From the geometrical relationships shown in Figure 1, the free energy change between the initial and final state can also be expressed as the ensemble average of the interaction energy difference at the midpoint between these states (denoted as $\langle V_B - V_A \rangle_{0.5}$ in Fig. 1). In this case, the free energy change, ΔG_2 , can be expressed as

$$\Delta G_2 = \langle V_H - V_P \rangle_{0.5}. \quad (8)$$

Here $\langle \dots \rangle_{0.5}$ denotes the ensemble average in the midpoint between the hydration state H and the point particle state P. This midpoint-sampling approach was used earlier.^{13,14} If states H and P are relatively similar, which is actually a prerequisite of the linear response approximation, the potential function describing the midpoint state can be expressed as $(1/2)(V_H + V_P)$ in the first-order approximation. Both V_H and V_P functions include two parts: the common interaction function V_C that describes the interactions between the water molecules themselves and the state-specific interaction functions V_H^a and V_P^a . The former function describes the interactions between the solute atom and the surrounding water in the state H whereas the latter describes the interactions between the point particle and water in the state P. V_P^a behaves as a hard-sphere interaction function that is characterized by a strong repulsive force when the interaction distance is less than r_0 and has a zero value when the distance is larger than r_0 . At room temperature the interwater distances are practically always larger than the value of r_0 (1.776 Å). To some extent, this assumption can be supported by the frequent estimate of the water radius as 1.4 Å in solvent accessible surface area calculations, which implies that the average distance between two water molecules is not less than 2.8 Å at room temperature. Therefore, it appears reasonable to substitute the potential function describing the midpoint state [i.e., $(1/2)(V_H + V_P)$] by $V_C + (1/2)V_H^a$ for all accessible interaction distances at room temperature. Similarly, $V_H - V_P$ in eq. (8) can be approximated by V_H^a . Thus, the free energy change in the second step can be calculated from the following equation:

$$\Delta G_2 = \langle V_H^a \rangle_{0.5}, \quad (9)$$

where $\langle \dots \rangle_{0.5}$ denotes the ensemble average in the midpoint state, which is described by the potential function $V_C + (1/2)V_H^a$. The value of $\langle V_H^a \rangle_{0.5}$ and, therefore, ΔG_2 can be easily obtained

from molecular dynamics or Monte Carlo simulations in this midpoint state.

In summary, the hydration free energy, ΔG , resulting from the introduction of a solute atom into the aqueous solution can be calculated as the sum of the free energy changes of these two steps, ΔG_1 [cf. eq. (7)] and ΔG_2 [cf. eq. (9)], as follows:

$$\Delta G = \Delta G_1 + \Delta G_2 = 1.49 k_B T + \langle V_H^a \rangle_{0.5}. \quad (10)$$

Computational Details

GENERAL DETAIL

The standard Sigma molecular dynamics package¹⁵ was used to perform all simulations reported in this article. Sigma uses the description of nonbonded interactions from the GROMOS force field¹⁶ and bonded interaction parameters that were developed independently.¹⁷ In this work the nonbond interaction parameters were taken from the OPLS force field model.^{18–21} The simulations were done with both the united-atom and all-atom force field models. The organic compounds tested in this work were the same as those used by Carlson and Jorgensen⁶ and included methane, ethane, methyl chloride, methanethiol, dimethyl ether, acetonitrile, methylamine, methanol, acetone, methyl acetate, acetic acid, acetamide, benzene, and pyridine. The atomic point charges for these compounds were reported earlier.^{18–21} They were obtained by either fitting to the experimental results in the united-atom model or fitting to the electrostatic potential surface (EPS) derived from *ab initio* calculations at the 6–31G* level in the all-atom model. In the course of all simulations, the internal degrees of freedom were not considered (i.e., all the molecules were fixed at their standard geometries based on the microwave structures).²² Thus, the possible contribution of the internal degrees of freedom was neglected. The SPC model was used to describe water.¹¹ All simulations were done at 298 K. The periodic boundary conditions were used, and the dimensions of the box of water were $20 \times 20 \times 20 \text{ \AA}^3$. The cutoff radius for the nonbond interactions was 10 \AA .

HYDRATION FREE ENERGY CALCULATIONS USING GLR METHOD

For the multiatomic compounds used in this study, we postulated that the contributions to the total free energy from individual atoms in eq. (10) were additive. Thus, the hydration free energies

for our data set were calculated from the following equation:

$$\Delta G = 1.49 n k_B T + \langle V_H^c \rangle_{0.5}, \quad (11)$$

where n is the number of atoms in the compound and V_H^c denotes the interaction function between the solvated compound and water molecules in the hydration state H.

After each compound was solvated with the SPC water, the whole system was equilibrated via 50 steps of conjugate gradient energy minimization followed by 60 ps of molecular dynamics simulation in the midpoint state, which was defined by the virtue of using $V_C + (1/2)V_H^c$ as a potential interaction function. The value of $\langle V_H^c \rangle_{0.5}$ was obtained from an additional 80-ps molecular dynamics simulation in the midpoint state. The statistical error was estimated as the average unsigned deviation of the hydration free energies calculated from the 10 consecutive 8-ps segments of the molecular dynamics trajectory during the whole 80-ps simulation.

HYDRATION FREE ENERGY CALCULATIONS USING CONVENTIONAL SG SIMULATION METHOD

We followed the SG protocol for the free energy simulation method in which the interactions of a solute compound with the water environment slowly vanish in the course of the simulations whereas the solute's structural integrity is maintained. The potential function of the system, V_λ , is made explicitly dependent on the coupling parameter λ :

$$V_\lambda = V_C + \lambda^m V_H^c. \quad (12)$$

As discussed above, V_C describes the interactions between the water molecules themselves and V_H^c describes the interactions between the hydrated compound and water. Nonlinear λ coupling was used to avoid the singularity problem in free energy simulations. Based on the Sigma default settings, $m = 3$ in eq. (12) was used for the attractive part of the van der Waals interaction function and the electrostatic interaction function and $m = 5$ was used for the repulsive part of the van der Waals interaction function. Thus, at $\lambda = 1$, V_λ describes the potential function of the system with the completely hydrated compound whereas at $\lambda = 0$ it describes the potential function of the system after the compound has completely disappeared from the aqueous solution. In the course of the molecular dynamics simulation λ slowly

changes from 1 to 0 (forward simulation). The free energy change in the course of this molecular replacement process (i.e., dehydration free energy ΔG_{dehyd}) is calculated by the thermodynamic integration method from the following equation:

$$\Delta G_{\text{dehyd}} = \int_{\lambda=1}^{\lambda=0} \langle \partial V_{\lambda} / \partial \lambda \rangle d\lambda. \quad (13)$$

In practice, this integral is approximated by the following formula that was used in our calculations:

$$\Delta G_{\text{dehyd}} = \sum_{\lambda=1}^{\lambda=0} (\partial V_{\lambda} / \partial \lambda) \delta \lambda. \quad (14)$$

To reduce the systematic error due to the hysteresis,²³ such a molecular replacement simulation was repeated in the reverse direction and the corresponding free energy change (i.e., hydration free energy, ΔG_{hyd}) was calculated from a similar equation:

$$\Delta G_{\text{hyd}} = \sum_{\lambda=0}^{\lambda=1} (\partial V_{\lambda} / \partial \lambda) \delta \lambda. \quad (15)$$

The final calculated hydration free energy value was obtained as the average of the results of the

forward and reverse calculations (i.e., $\Delta G = (1/2) \cdot (\Delta G_{\text{hyd}} - \Delta G_{\text{dehyd}})$).

After an initial 60-ps molecular dynamics equilibration of the whole system with the solute compound and water solvent, 400-ps forward molecular replacement simulation was done for each compound (i.e., λ was changed from 1 to 0). The system without the hydrated organic compound was then dynamically equilibrated for 60 ps at $\lambda = 0$, followed by 40-ps reverse molecular replacement simulation (i.e., λ was changed from 0 to 1). The statistical error was estimated using the hysteresis because the former was shown to be proportional to the latter.²³

Results

HYDRATION FREE ENERGY CALCULATIONS USING ALL-ATOM FORCE FIELD

This set of simulations was conducted using the van der Waals parameters from the OPLS force field and ESP charges derived by the Jorgensen group.^{18–21} The results of these calculations by the GLR method and the conventional SG method are both presented in Table I. The mean unsigned deviation of the hydration free energies for all 13

TABLE I. Comparison of Hydration Free Energies (kcal / mol) Obtained Experimentally and Calculated Using GLR and SG Methods with OPLS All-Atom Force Field.

Compounds	$\langle V_{\text{H}}^{\text{c}} \rangle_{0.5}$	1.49 $nK_{\text{B}}T$	ΔG_{hyd}		Exp ^c
			GLR ^a	SG ^b	
CH ₃ OH	−10.33	5.29	−5.0 (0.9)	−4.8 (0.8)	−5.1
CH ₃ NH ₂	−10.76	6.17	−4.6 (0.4)	−4.2 (0.6)	−4.6
CH ₃ CN	−10.93	5.29	−5.6 (0.4)	−4.5 (0.3)	−3.9
CH ₃ OCH ₃	−9.70	7.94	−1.8 (0.4)	−2.1 (0.2)	−1.9
CH ₃ SH	−6.75	5.29	−1.5 (0.4)	−0.1 (0.7)	−1.2
CH ₃ Cl	−5.79	4.41	−1.4 (0.3)	−1.0 (0.2)	−0.5
C ₂ H ₆	−4.12	7.06	2.9 (0.3)	2.8 (0.3)	1.8
CH ₃ CONH ₂	−20.11	7.94	−12.2 (1.0)	−12.9 (0.3)	−9.7
CH ₃ COOH	−15.70	7.06	−8.6 (0.6)	−8.5 (0.3)	−6.7
(CH ₃) ₂ CO	−13.33	8.82	−4.5 (0.8)	−4.3 (0.6)	−3.8
CH ₃ COOCH ₃	−15.17	9.70	−5.5 (0.6)	−5.1 (0.3)	−3.3
C ₆ H ₆	−11.14	10.58	−0.6 (0.3)	−0.9 (0.1)	−0.8
C ₅ H ₅ N	−14.52	9.70	−4.8 (0.4)	−4.7 (0.1)	−4.7
Average error ^d			0.9	0.9	

^a The numbers in the parentheses are the mean unsigned deviations obtained from 10 consecutive 8-ps fragments of the total 80-ps molecular dynamics trajectory.

^b The numbers in the parentheses show the hysteresis, which is calculated as the sum of the free energy values obtained from the forward and reverse molecular replacement simulations.

^c Reference 6.

^d Mean unsigned deviation from the experimental data.

organic compounds between experimental data and those calculated by the GLR method is 0.9 kcal/mol. The mean unsigned deviation between the values calculated by the SG method and those obtained from the experiment is also 0.9 kcal/mol. The mean unsigned deviation between the two series of calculated values is only 0.4 kcal/mol. Thus, the two sets of calculated data strongly correlate with each other ($R^2 = 0.93$) as illustrated by Figure 3.

Seventy-eight values of the relative hydration free energy, $\Delta\Delta G_{\text{hyd}}$, for all pairs of these 13 compounds were also calculated from absolute hydration free energies shown in Table I. The mean unsigned deviations from experimental data were 1.3 kcal/mol for the GLR and SG methods. The mean unsigned deviation between the two series of calculated data was 0.6 kcal/mol.

HYDRATION FREE ENERGY CALCULATIONS USING UNITED-ATOM FORCE FIELD

We repeated the simulations of almost the same set of compounds using the united-atom OPLS model that provides standard OPLS parameters for van der Waals and electrostatic interactions. The results calculated by the GLR method and the SG method are both presented in Table II. The mean unsigned deviation between the hydration free energies of 13 organic compounds calculated

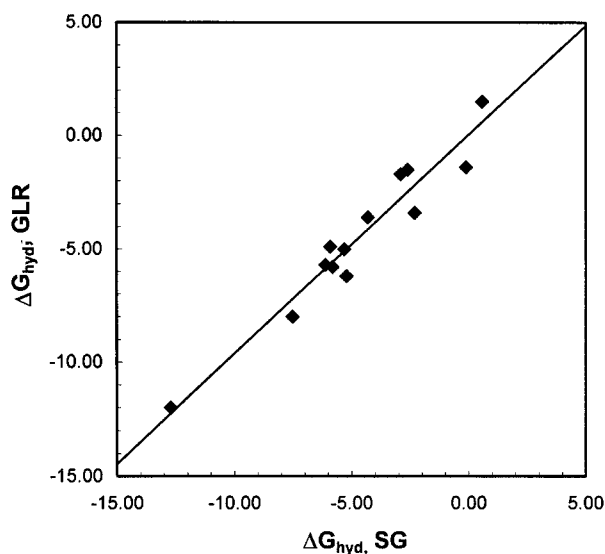


FIGURE 3. Correlation between the hydration free energies (kcal/mol) of 13 organic compounds (see Table I) calculated from the GLR and SG methods using the all-atom OPLS force field.

by the GLR method and those obtained from the experiment is 1.0 kcal/mol. The mean unsigned deviation between the values calculated by the SG method and those obtained from the experiment is 1.1 kcal/mol. The mean unsigned deviation between the two series of calculated values is 0.5 kcal/mol (i.e., almost the same as that in the all-atom force field model). Once again we observed a strong correlation between the two sets of calculated data ($R^2 = 0.96$) as shown in Figure 4.

Seventy-eight values of relative hydration free energies were also calculated for all pairs of compounds using the absolute hydration free energies reported in Table II. The mean unsigned deviations from the experiment were 1.1 kcal/mol for the GLR method and 1.2 kcal/mol for the SG method. The mean unsigned deviation between the two series of calculated data was 0.8 kcal/mol.

Discussion

We have described a novel linear response method that, to the best of our knowledge, is the first application of a uniform treatment of van der Waals and electrostatic interactions in the free energy calculations by the linear response approximation. The main assumption of this method is that the hydration process can be divided into two steps. In the first step a point particle is introduced into water, which, according to the scaled particle theory, creates a cavity of the water size. For the SPC water model the free energy change of this step can be estimated from the scaled particle theory as $1.49k_B T$. In the second step this point particle is transformed into the solute atom and the accompanying free energy change can be calculated by the linear response method. For the n -atomic molecule we assume the additivity of the atomic hydration free energies (i.e., the free energy of the first step above is calculated as $1.49nk_B T$). The resulting equation for the hydration free energy calculations [eq. (11)] does not contain any empirically determined parameters and therefore should be easily applied to various systems. The coefficient (1.49) in eq. (11) was derived using $r_0 = 1.776 \text{ \AA}$, which is specific for the SPC water model. For other water models the value of r_0 should be calculated independently, which may result in a different value of the coefficient in eq. (11). Thus, the GLR method can be used for any solvent model for which the basic assumptions of the linear response theory are valid.

TABLE II.

Comparison of the Hydration Free Energies (kcal / mol) Obtained Experimentally and Calculated Using GLR and SG Methods with OPLS United-Atom Force Field.

Compound	$\langle V_H^c \rangle_{0.5}$	$1.49nK_B T$	ΔG_{hyd}		Exp ^c
			GLR ^a	SG ^b	
CH ₃ OH	-9.29	5.29	-4.0 (0.6)	-4.0 (0.2)	-5.1
CH ₃ NH ₂	-9.45	6.17	-3.3 (0.5)	-4.5 (0.3)	-4.6
CH ₃ CN	-9.25	5.29	-4.0 (0.6)	-2.6 (0.4)	-3.9
CH ₃ OCH ₃	-7.94	7.94	-0.0 (0.7)	-0.4 (0.1)	-1.9
CH ₃ SH	-7.16	5.29	-1.9 (0.5)	-0.9 (0.5)	-1.2
C ₂ H ₆	-4.04	7.06	3.0 (0.4)	2.8 (0.0)	1.8
CH ₄	-1.92	4.41	2.5 (0.3)	2.6 (0.2)	2.0
CH ₃ CONH ₂	-17.66	7.94	-9.7 (0.9)	-10.6 (0.1)	-9.7
CH ₃ COOH	-12.53	7.06	-5.5 (0.9)	-5.5 (0.4)	-6.7
(CH ₃) ₂ CO	-10.39	8.82	-1.6 (0.5)	-1.7 (0.2)	-3.8
CH ₃ COOCH ₃	-10.52	9.70	-0.8 (0.4)	-0.7 (1.0)	-3.3
C ₆ H ₆	-11.13	10.58	-0.6 (0.4)	-0.4 (0.8)	-0.8
C ₅ H ₅ N	-14.03	9.70	-4.3 (0.4)	-5.5 (0.8)	-4.7
Average error ^d			1.0	1.1	

^a The numbers in the parentheses are the mean unsigned deviations obtained from 10 consecutive 8-ps fragments of the total 80-ps molecular dynamics trajectory.

^b The numbers in the parentheses show the hysteresis, which is calculated as the sum of the free energy values obtained from the forward and reverse molecular replacement simulations.

^c Reference 6.

^d Mean unsigned deviation from the experimental data.

The accuracy and efficiency of the GLR method was tested by calculating the hydration free energies of 14 diverse organic compounds considered earlier by Carlson and Jorgensen.⁶ In our calculations we used almost the same set of force field

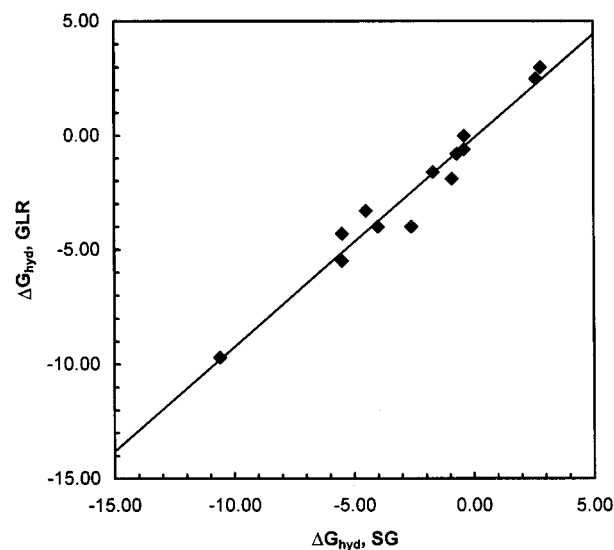


FIGURE 4. Correlation between the hydration free energies (kcal / mol) of 13 organic compounds (see Table II) calculated from the GLR and SG methods using the united-atom OPLS force field.

parameters and repeated the calculations with both the united-atom and all-atom force field models. The mean unsigned deviation between experimental and calculated hydration free energies was about 1.0 kcal/mol with the united-atom model and 0.9 kcal/mol with the all-atom model. Taking into account that the random statistical errors of the hydration free energies calculated by the GLR method were around 0.5 kcal/mol (cf. Tables I, II), these mean unsigned deviation values indicate that the GLR method can provide a good estimate of the experimental hydration free energies. The agreement between the two sets of calculated data in our work is especially remarkable (cf. Figs. 3, 4) with the mean unsigned deviation of 0.4 kcal/mol with the all-atom model and 0.5 kcal/mol with the united-atom model. The strong correlation between the GLR and SG calculations further indicate the robustness of this novel linear response method, because the errors that might originate from the choice of the force field model have been excluded. The results with the GLR method were obtained from the 80-ps molecular dynamics trajectory, whereas the SG free energy calculations required at least 800 ps of molecular dynamics sampling (400 ps of the forward and 400 ps of the reverse molecular replacement simulations, not counting the 60-ps equilibration). Thus, the GLR

method as implemented in this work is about one order of magnitude more computationally efficient as compared to the SG method.

To compare our results with those of Carlson and Jorgensen⁶ and Jorgenson and Tirado-Rives,²⁴ the RMS deviations between the GLR data and the experimental data were also calculated. These RMS values were 1.2 kcal/mol for the all-atom model and 1.3 kcal/mol for the united-atom model, respectively. These values are slightly higher than the RMS deviations of 0.8 kcal/mol for the all-atom model and 0.9 kcal/mol for the united-atom model, respectively, as reported by Carlson and Jorgensen.⁶ Although our results are slightly worse than those of Carlson and Jorgensen,⁶ this is not particularly surprising because the GLR parameters were derived theoretically while Carlson and Jorgensen⁶ used three adjustable parameters that were determined by fitting to the experimental data.

To address the performance of the GLR method for the relative hydration free energy calculation as compared to Carlson and Jorgensen calculations,⁶ 78 values of the difference between ΔG_{hyd} for all pairs of compounds (cf. Tables I, II) were also calculated. The accuracy of the relative free energy calculations was slightly worse than that for the absolute free energy calculation. The RMS deviation values between 78 experimental and calculated relative hydration free energies were 1.5 kcal/mol (all-atom model) and 1.3 kcal/mol (united-atom model), whereas the RMS deviation values between 13 experimental and calculated absolute hydration free energies were 1.2 kcal/mol (all-atom model) and 1.3 kcal/mol (united-atom model). The relatively higher values of the RMS deviations of relative versus absolute hydration free energies are due to the fact that the relative hydration free energies are calculated as the difference of absolute hydration free energies for two respective compounds. In fact, the SG method produces RMS deviations that are even slightly higher than those from the GLR method: 1.7 kcal/mol (all-atom model) and 1.4 kcal/mol (united-atom model) for relative hydration free energies compared to kcal/mol (all-atom model) and 1.3 kcal/mol (united-atom model) for absolute hydration free energies. The same trend was also observed with Carlson and Jorgensen's data.⁶ Using their absolute hydration free energies calculated with fitted parameters in their equation, we obtain RMS deviations of 78 values of relative hydration free energy from experimental data of 1.1 kcal/mol (all-atom model) and 1.3 kcal/mol

(united-atom model). The RMS deviations of 13 values of the absolute free energy from experimental data are 0.8 kcal/mol (all-atom model) and 0.9 kcal/mol (united-atom model).

Obviously, the inclusion of the cavity contribution term is crucial for the accurate calculations of the hydration free energies by the linear response approximation. Carlson and Jorgensen also reached the same conclusion in their work, except that they used molecular surface area, solvent accessible surface area, or molecular volume to describe this contribution term.⁶ According to the GLR model presented here, the inclusion of the cavity contribution term is crucial to prevent the singularity effect and make the next step of the hydration process adequately described by the linear response approximation.

For the larger compounds, such as polypeptides, or in the more heterogeneous environment (e.g., receptor binding site) more careful consideration may be needed to find a correct expression to reflect the cavity contribution. However, in practice we are often interested in the calculation of relative free energy changes (e.g., relative binding free energy) and the possible systematic errors from cavity terms can often cancel each other in the parallel processes of a thermodynamic cycle.²⁵ Thus, the GLR method may be applied to the relative binding free energy calculations as well. We are currently investigating this application of the GLR method using several HIV protease-1 inhibitor complexes. Another possible challenge in future studies would be to account for the contribution of the internal degrees of freedom, which was neglected in the present work. These future studies will determine the actual ranges of the applicability and reliability of the GLR method.

Acknowledgments

The authors wish to acknowledge Drs. J. Hermans, T. O'Connell, and I. Vaisman for helpful critical suggestions and Dr. Julian Tirado-Rives for providing the OPLS parameters for compounds used in this work.

References

1. Karplus, M.; Petsko, G. A. *Nature* 1990, 347, 631.
2. Kollman, P. A. *Chem Rev* 1993, 93, 2395.

3. Van Gunsteren, W. F.; Berendsen, H. J. C. *Angew Chem Int Ed Engl* 1990, 29, 992.
4. Jorgensen, W. L. *Acc Chem Res* 1989, 22, 184.
5. Aqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng* 1994, 7, 385.
6. Carlson, H. A.; Jorgensen, W. L. *J Phys Chem* 1995, 99, 10667.
7. Marcus, R. A. *Annu Rev Phys Chem* 1964, 15, 155.
8. McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. *J Phys Org Chem* 1997, 10, 563.
9. Pierotti, R. A. *Chem Rev* 1976, 76, 717.
10. Pierotti, R. A. *J Phys Chem* 1965, 69, 281.
11. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, Holland, 1981.
12. Weast, R. C. *CRC Handbook of Chemistry and Physics*; CRC Press: Cleveland, OH, 1977.
13. Dejaegere, A.; Karplus, M. *J Phys Chem* 1996, 100, 11148.
14. King, G.; Barford, R. A. *J Phys Chem* 1993, 97, 8798.
15. Hermans, J. *Sigma Manual*; University of North Carolina: Chapel Hill, NC, 1997.
16. Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* 1984, 23, 1513.
17. Ferro, D. R.; McQueen, J. E.; McCown, J. T.; Hermans, J. *J Mol Biol* 1980, 136, 1.
18. Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J Comput Chem* 1993, 14, 1240.
19. Jorgensen, W. L.; Nguyen, T. B. *J Comput Chem* 1993, 14, 195.
20. Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J Phys Chem* 1994, 98, 13077.
21. Jorgensen, W. L.; Severance, D. L. *J Am Chem Soc* 1990, 112, 4768.
22. Harmony, M. D.; Laurie, V. W.; Kuczowski, R. L.; Schwen-denman, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G. *J Phys Chem Ref Data* 1979, 8, 619.
23. Hermans, J. *J Phys Chem* 1991, 95, 9029.
24. Jorgensen, W. L.; Tirado-Rives, J. *Perspect Drug Discov Design* 1995, 3, 123.
25. Van Gunsteren, W. F.; Weiner, P. K. *Computer Simulation of Biomolecular Systems*; Escom: Leiden, The Netherlands, 1989.